ALL-OPTICAL WDM PACKET NETWORKS

S. P. Monacos, J. M. Morookian, L. J. Davis,
L. A. Bergman, S. Forouhar, J. R. Sauer

April 17, 1995

ABSTRACT

This project is to develop the components and subsystems for the implementation of a multi-GHz optoelectronic data transport network using self-routing packets in a multi-hop network. The short packet payloads are compressed using optical wavelength division multiplexing techniques, and remain optical from source to destination while traversing the switching nodes. The routing is done with a lean, self-routing *hot potato* protocol in order to avoid the need for data storage at the switching nodes and to provide a fixed node latency equivalent to a few meters of fiber. Sustainable throughput both in to and out of the electronic host at each node should exceed 10 gigabit/see. Some technical details of the switching nodes and interfaces of tile recirculating shuffle network, and the stepped wavelength laser arrays and testbed will be given.

*This work was supported by the 13 MDO/IST.

rate communication architectures, supercomputer links, separation of control signals from data signals, phased arrays, and concurrent (parallel) processors. This multi-wavelength format can dramatically increase the capacity of present transmission systems without requiring significant technological developments in the modulation bandwidth of high speed electronics, transmitters and receivers.

Current lightwave communications systems are limited by the use of electronics for signal regeneration, packet routing and buffering at intermediate switching nodes. If a WDM data format is used with electronics in the data path, the packets must be optical ly demultiplexed and converted into separate electronic channels. Each optical channel requires separate hardware to buffer and route the data. Finally, the individual channels are converted back to optical form and multiplexed back into one fiber. By avoiding the use of optoelectronic conversions at intermediate nodes, WDM techniques can further leverage today's technology of intensity–modulated, direct detection systems to realize much higher capacity networks without modification of the network fabric due to the WDM format.

The components needed to realize the full benefit of a WDM network are monolithic WDM laser diode and detector arrays, a WDM network interface and an all-optical data path switching node. In this paper, wc describe on going work in these areas. Section 2 discusses a ncw type of recirculating network topology well suited to routing of optical packets. In this section, we further describe a switching node implementation for this network and a conceptual interface design to interconnect this network to existing electronic networks. In section 3, we present a simplified version of this interface used to construct a 4-channel WDM HIPPI link testbed. In this section, we also discuss progress on a 4-element single chip stepped wavelength DFB laser diode array for high–data rate WDM communication systems. Section 4 discusses potential applications for multi-gigabit/sec communications networks. Future work is presented in section 5, and conclusions are given in section 6.

2. Multi-Cylinder ShuffleNet (MCSN)

To take full advantage of the benefits of WDM technology', we desire a network capable of routing optical packets. The fundamental problem in routing optical packets is that there is currently no good way of storing data in optical form to handle contention problems. As such, a new topology and control structure was developed which places no timing constraints on arrival times of data packets at switching nodes and fixed set up latencies for making routing decisions at the switching nodes of the network. The end result is the MCSN architecture with a short packet deflection protocol (hot potato) [1] which can bc easily interfaced to one or more industry standard protocols for connection to existing networks.

The goal of the MCSN is to provide a methodology well suited to routing of optical WDM packets without header modification or optoelectronic conversions and with very low routing delay for fine-grain distributed supercomputing. Detailed simulations of this network exhibit crossbar like routing characteristics with near 100% availability of the network input ports [1]. Current work focuses on developing a proof-of-concept all-electronic MCSN prototype switching node as a precursor to building an electronic 8-node MCSN. In this section we describe the basic architecture of the MCSN and the functional requirements for the switching nodes and interfaces for this network.

2.1. MCSN Architecture

The MCSN topology described in Ref. [1] is a network which uses a recirculating shuffle network (SN) with an all-optical data path from source to destination. The difficulty with this network architecture is that packet deflections result even at very low network loading [1]. To improve network performance wc augment the basic SN topology with multiple parallel copies of the original topology called *routing cylinders.* Additionally, the switching nodes and links of the network dynamically store packets and provide congestion control within the network by routing blocked packets onto alternative routing cylinders. The architecture is easily scalable and uses the hot potato protocol in Ref. [3] but without age/priority information. The MCSN architecture is also designed for *packet asynchronous* traffic, thus avoiding the need to synchronize packets entering the network.

Figure 1 is an example of an 8-node SN topology. An R-cylinder MCSN topology is topologically equivalent to the generic SN topology but has R parallel perfect shuffle interconnections between stages of SN nodes (i.e. wc expand each node–to–node link to *R data paths* in the R-cylinder SN) [1]. We accommodate this augmentation to the generic SN topology by expanding the number of ports at each node to R times that of the basic SN switching node shown in Fig. 2. Additionally, some of the R data paths to the host may instead be used for local recirculation links of blocked packets [1].

2.2. MCSN Node

The MCSN switching nodes internally use a *Permutation Engine (PE),* which is described in detail in Ref. [2]. The PE is a simple distributed routing control mechanism for routing packet-asynchronous data using only local traffic information to make routing decisions. For the R-cylinder configuration, the PE switching nodes also provide dynamic routing of packets between SN cylinders.

Figure 3 is a block diagram representation of a MCSN switching r-rode. This node consists of the header detection logic, the header translation stage, the header plane, the first-in-first-out (FIFO) buffers, and the data plane. The header detection logic is used to detect an n bit-serial packet header and convert it to an n-bit parallel header. The translation stage is used to convert a global packet header to a local node output port number. The header plane routes packet headers to establish the input port to output port connection based on the PE routing algorithm. The FIFO buffers are used to delay incoming packets for a fixed time interval based on the set up times of the header detection logic, the translation stage, and the header plane. Finally, the data plane is used to route packets based on the control signals from the header plane.

2.2.1 Protot ype Switching Node

A prototype all-electronic implementation of the switching node in Fig. 3 is currently in development at JPL. The basic configuration is a 12-input to 12-output PE on a 6Ux220mm VME wire wrap card, The number of input/output ports was selected based on the MCSN design requirements for an 8-node system [1]. The header translation stage shown in Fig. 3 was omitted to simplify testing and verification of the node design. For an 8-node MCSN demonstration system, a small amount of translation logic can be incorporated into the header detection logic of each node. This logic must be customized for each node to implement the deflection routing protocol similar to that in [3] but without age/priority information. For large scale MCSNS (> 100 nodes) a simple memory look up table an be used to implement the translation stage. The layout for this prototype node is shown in

Fig. 7. In addition to the basic node design shown in Fig. 3, Fig. 7 also shows a packet generator. This component is used to generate packet headers for injection into the node to verify routing of the headers and asses packet flow through the node.

2.3. MCSN Network Interface

The serial packet format shown at the top of Fig. 4 is suitable for packet switched networks which can handle long duration packets. The performance of such networks is significantly enhanced by using buffers to avoid packet deflection from output port contention [4], [5]. For the proposed all optical data path switching node, such buffering capability is not currently available. In order to reduce the probability of packet contention at the output ports of a switching node, it is desirable to reduce the packet duration by parallelizing the header, data and trailer information as shown in Fig. 4. For an all–optical data path network, this scheme is realized by using a wavelength division multiplexed (WDM) format where the start/stop bits, header, *m* data words and clock are encoded onto different optical frequencies.

The basic problem is that existing protocols are structured to utilize the strengths of electronics, namely good component functionality with data rates in the 100's of megabits/sec range, static storage and μs path configuration times. These properties have resulted in a variety of circuit switched architectures where the overhead for path set up is reduced by sending long packets. This construct is similar to a burst transfer meaning a large block of data is sent once the bus – data path – has been acquired, The problem with this approach is that the path set up time grows to 10's of rns for WANS and is unacceptable for low routing delay applications.

An alternative scheme is to build a datagram or packet switched network. In this protocol a header is prefixed to a packet and is used to dynamically route a packet through the network. In the electronic domain such schemes utilize a store–and- forward approach with electronic buffers for packet storage until the data path of a switching node can be established. This type of network is better suited to WANS because the path set up is done locally at each switching node so that a global request/acknowledge cycle is avoided.

2.3.1. Optoelectronic Space Time-Multiplexer

In looking at the requirements for an all–optical data path network and comparing them to the operation of existing electronic networks, we see a dichotomy in terms of the required functionality of each type of network. The all–optical data path network requires switching nodes capable of handling small asynchronous packets with fixed routing latency. The hot potato deflection protocol provides this functionality by misrouting/deflecting a packet instead of storing it in a buffer for an arbitrary period of time until the desired routing path is achieved. Host devices connected to this network, however, use industry standard protocols such as HIPPI to transmit information. Protocols such as HIPPI are store–and--forward in nature and allow a header or packet to sit at an intermediate switching node for arbitrary time periods as a result.

To connect host devices using industry standard protocols to an all–optical data path network requires developing an interface which converts from existing protocols well suited to electronics to a hot potato style scheme well suited to optics. Such an interface was conceptualized and is called a *Space--Time Multiplexer* (STM) due to its ability to multiplex *N* multiple electronic sources in a time division multiplexed (TDM) Fashion onto a multi--wire link. The space multiplexing function occurs by mapping one electronic source onto an *m*-bit wide link, Figures 5 and 6 show the transmitter and receiver sections of this interface respectively to realize an *m*-bit wide link, where each bit corresponds to onc wavelength in the WDM data format in Fig. 4.

The three functional elements of the STM interface are segmentation of a large packets into small cells, flow control, and header insertion/detection. "l'he first element looks to break up a large packet into many smaller packets. Many electronic protocols allow for large packets to reduce the overhead incurred in setting up the data path. For the envisioned optical data path network, however, large packets pose a serious blocking problem in that optical packets can not be statically stored and will be misrouted if a given packet ties up a routing path for an extended period of time. Thus, the efficiency of a deflection routing network is improved by deceasing the packet size. The third element is a consequence of this packetizing process and entails inserting a header before each packet at the transmitter side and removing this header at the receiver. The second element addresses the global request/acknowledge issue by using local flow control between the local host device and the STM interface. This scheme allows for source throttling as an indirect global flow control mechanism without incurring the long time of flight penalty for a global path set up.

3. WDM Network Components

In this section we describe recent progress in developing the necessary components needed to build the MCSN. These components consist of a 4-wavelength WDM HIPPI link and 4-element stepped wavelength laser arrays under development at the Micro Device Laboratory (MDL). Future directions concerning these components are presented in section 5.

3.1. WDM HIPPI Link

A four channel, multi–gigabit/s, full duplex WDM HIPPI extender link is currently in development at JPL. A block diagram of one of the two identical interfaces is shown in Fig. 8. HIPPI devices at one end of the link provide four parallel HIPPI inputs (32 bits wide x 25 MHz= 800 megabit/see each) which are routed to the transmit side of the} HIPPI parallel–serial printed circuit board, which multiplexes the parallel data clown to a serial stream at a 1.2 gigabit/see. The four serial HIPPI lines arc used to modulate a four--e lement distributed feedback (DFB) laser diode array which was designed and built at JPL. The optical output from the DFB array is coupled to an array of four optical fibers which are combined to a single optical fiber using a fused coupler. The single fiber output is routed to a 10km optical fiber link. At the receive end, a 1:4 fused coupler provides four identical copies of the WDM optical signal to a bank of tunable optical bandpass filters, Each filter is manually tuned to pass a single WDM channel and reject the others. Four commercial fiber optic receivers provide ECL-compatible outputs corresponding to the four 1.2 gigabit/see serial HIPPI streams. The receiver section of the HIPPI parallel-serial board demultiplexes the serial streams back to 32 bit–wide HIPPI, which is returned to the destination ports on the HIPPI devices.

3.2. 4-element WDM Laser Diode Array

The core of the system described above is a monolithic laser diode array developed at JPL's Micro Devices Lab (MDL). The transmitters have four side-by-side single mode DFB lasers made on a single substrate, with each laser emitting light at a slightly different wavelength in the 1,55 μm region. The laser design is an InP–based ridge waveguide laser. Due to the simplicity of fabrication and less stringent fabrication tolerances compared to buried heterostructure lasers, ridge waveguide lasers are seen to have a strong potential for commercial use[6].

3.2.1. Laser Array Growth and Fabrication

The laser wafers were prepared by atmospheric pressure metal--organic chemical vapor deposition(MOCVD) on (100)–oriented n^+InP substrates. The active region consists of 4 compressively strained (ε=1%) InGaAsP quantum wells, each 94 Å wide, with 150 Å barriers of InGaAsP (λ=1.2 pm). The optical confinement is provided by a stepped separate confinement heterostructure(SCH) region consisting of 900 Å InGaAsP (λ=1.2 μm) and 800 Å InGaAsP (λ=1. 15 pm), with InP as the top and bottom cladding material. The conduction band profile of the complete laser structure is shown in Fig. 9. Broad area lasers were fabricated to evaluate the quality of the material; measurement of the threshold current and slope efficiency versus cavity length allowed the extraction of the internal quantum efficiency (60 %) and the internal loss (17.4 /cm).

Fabrication of this material into 4-clement DFB laser arrays requires e---beam writing of the diffraction gratings, an MOCVD regrowth, and the fabrication of the ridge waveguide structure. The top 4 layers of the laser structure (contact, 2 InP layers, and etch stop in Fig. 9) are removed in order to define the distributed feedback grating in the SCH region. The pitch of the grating for the individual lasers is determined by the modal index and the design criteria of four wavelengths in the range from 1.54-1.56 μm (to be compatible with erbium doped fiber amplifiers). This leads to four grating pitches in the range from 2375--2400 Å. The gratings are e–beam defined in PMMA, and etched into the InGaAsP (1.15 μm) layer using an aqueous solution of HBr and HNO_3. MOCVD is then used to regrow the same 4 layers back onto the structure. Ridge waveguide lasers are then fabricated from this regrown structure. First, the *p* contact (Ti/Pt/Au) is deposited and annealed; each contact is nominally 3.5 μm wide. A self---aligned wet chemical etch is used to define the ridge waveguide structure. Use of an etch stop allows for reproducible waveguide definition with a pre–determined amount of index—guiding. The amount of index—guiding is dictated by the InP spacer thickness. After the ridge definition etch, polyimide is applied to the wafer and then cured. Oxygen–based reactive ion etching (RIE) is then used to open the polyimide to the p contact. The final top side processing is the lithography and evaporation for the contact metal (Cr/Au). The wafer is then lapped to a thickness of-100 μm, and then a back contact metal is evaporated (AuGe/Ni/Au). A final anneal completes the laser fabrication, and the devices are then scribed and cleaved.

The devices arc soldered to a silicon submount as shown in Fig. 10 and run CW. The submount can be fit into a variety of packages and the laser spacing (250 pm) is designed to be compatible with silicon v–groove based fiber arrays. The light vs. current characteristics of a 300 μm long, 4-element laser array is shown in Fig. 11 (a), showing the uniformity of the threshold and slope efficiency of the devices. Fig 11 (b) shows the spectral characteristics of this same array for a drive current of 50 mA, displaying a side mode suppression ratio greater than 20 dB. The finished laser arrays have wavelength separations of approximately 5 nm, very uniform threshold currents as low as 15 mA, output power of several mW, and excellent sidemode suppression ratios.

An important aspect of the WDM laser arrays is the reproducibility of the absolute wavelength and the wavelength spacing. Implementation of WDM systems requires a wavelength reference and definition of the required wavelength spacing in order to build the proper demultiplexing components, However, one finds that the absolute wavelength of the laser emission is directly proportional to the modal index, which can be affected by a number of process variations. Seemingly minor variations in the ridge width and etch depth (less than 1000 Å) can significantly affect the emission wavelength of a DFB laser. It is interesting to note that ridge waveguide structures arc less affected by such processing variations than buried heterostructure devices. Fig ure 12 shows a calculation of the change in emission wavelength with variations in the ridge width – for a ridge width of 3.5 μm, a 1000 Å in the ridge width will change the emission wavelength by approximately 0.5 Å. Buried heterostructure lasers show a much larger variation in the emission wavelength with changes in the active region width [7]. The flexibility of the ridge waveguide structure with respect to process variations allows for good control over the wavelength emission of the laser array. Figure 13 shows the emision wavelength of the different elements in several laser arrays, displaying the high degree of uniformity in the absolute wavelength and wavelength spacing achieved in our devices. The output wavelength spacing is approximately 5 nm, with a variation ± 1 nm. This accuracy should be adequate for the first generation of WDM devices; however, specifications for future WDM systems may require wavelength spacing of 0.8 nm (100 GHz) [7].

3.2.2. Device Speed Performance Tests

Before integration into tbc WDM I lIPPI extender, individual laser arrays are tested to insure that they meet the bandwidth requirement for use in a gigabit/see system. First, the laser diodes are modulated with a sharp electrical step function. A block diagram of the test setup is shown in Fig. 14. A time-domain relflectometer (TDR) plug--in module for the Tektronix CSA-803 oscilloscope provides a short electrical step function (rise time <100ps). This signal is fed from the instrument via coaxial cable to a test jig in which the laser array has been mounted. 'l'he 'bias–T' circuit shown in this figure is built using a microstrip design incorporating leadless chip components. The light output is directed to a high speed analog detector which converts the optical response to an electrical signal which is fed back to the CSA-803 oscilloscope for viewing.

The optical output generally exhibits a degree of cyclical overshoot and undershoot called the relaxation oscillation which gradually dies down as shown in Fig. 15. The frequency of the oscillation places an upper bound on the modulation rate possible under particular bias and drive conditions. For frequencies beyond the relaxation oscillation, the laser response drops off sharply. Therefore, the frequency content of the modulating signal should be kept below the relaxation resonant frequency [8]. It is also preferable for the damping of the oscillations to be high. As the current bias of a laser diode is increased in the linear regime of the L-I curve, the resonant relaxation frequency and the degree of damping increase. In the oscillogram of Fig. 15 the pronounced oscillation is due to the current bias being close to the threshold current. Figure 16 plots the relaxation resonance against the current bias. It can be seen that this oscillation approaches 3GHz at higher bias currents. It should be noted that the results of this test are affected by parasitic in both the laser module itself and the surrounding test setup. Thus, the observed resonance is likely limited by the chip components, cables, connectors, and other elements of the test apparatus, but it does indicate that modulation up to the desired rate of 1.2 gigabit/sec is possible.

In a second test, the rise and fall times produced in the optical waveform are observed by using a square wave input to each laser. The sum of these should be less than the period of one bit for the desired modulation rate, which is approximately 800 ps for serial HIPPI rates. In Fig. 17, the rise and fall times are limited by the speed of the electrical input to the laser diode (and not the laser response itself), which in this case is a high–speed silicon ECL–level driver.

3.3 WDM Link Components and System Testing

After completing the testing of individual laser array elements, the next step is to integrate the devices into the WDM HIPPI link for testing and performance characterization of the system. In this section, we describe the details of interfacing the laser array to the driver electronics and coupling of the laser elements to V-groove fiber array. Details concerning system testing are also presented.

3.3.1 Laser Diode Module

The DFB array is mounted on a block measuring roughly 250x 300x 100 roils. As shown in Fig. 18 the submount is bolted to a custom aluminum block. To the top of this block is mounted a microstrip design printed circuit board on a PTFE substrate which contains the required bias TEE circuitry. The signal input to the bias TEE is routed to the edge of the board. The connector tabs are soldered to these traces and the bodies of the connectors are bolted to the aluminum block for mechanical stability. Mini-clip type connectors are used to bring the bias currents to the inductors in the DC path of the bias TEE.

The optical fiber array to which the light output of the DFBs are coupled, consists of four optical fibers cemented between two silicon v–grooves with an inter-–groove spacing of 250 um to match the physical DFB spacing. The front surface of the fiber array is polished flush to the v–groove endface. The fiber v–groove assembly is positioned in front of the DFB array using a precision 6–axis translation/tilt stage. Both the aluminum block and translation stage may be bolted to an optical bench or to the bottom of a more portable chassis.

Additional printed circuit boards mounted in a card cage along with the HIPPI parallel-serial boards provide a stable bias current to the DFB elements. The bias circuitry provides from O to 80 mA of bias current and incorporates a 'slow start' feature which gradually ramps up the current to eliminate transient effects. The same printed circuit boards also contain the fiber optic receivers and related circuity.

3.3.2 **HIPPI Protocol Operation**

A HIPPI Tester, made by Input Output Systems Corporation, was used to send and receive parallel HIPPI data in the testbed. Figure 19 shows the connections between a source and destination device connected by a HIPPI interface. It performs a number of signaling tests and data loopback tests which provide a measurement of the word error rate of the link. Various types of data (all 1 's, all O's, walking 1 's, pseudo--random, etc.) may be sent under various signaling conditions.

The following discussion of the HIPPI specification provides background for later discussions on link operation [9], [10], There are 32 data lines, 4 lines of parity, CLOCK, several handshake lines (REQUEST, CONNECT, READY, PACKET, and BURST), and two interconnect lines (source–to–destination and destination--to–source). The two interconnect lines verify the 'hard' connection between the source and destination. If one or more of the interconnect lines are false, then either the HIPPI cables are not connected or one of the HIPPI devices is not powered on. No other signaling is allowed until both interconnect lines are true

The operation of the remaining control lines during a typical data transfer is shown in Fig. 20. When the source desires to make a connection, it will assert the REQUEST line and place the l–field (which contains information to select the desired destination) on the 32–bit data bus. The destination will respond with connect to make the connection. At this point, the source may assert PACKET to indicate a packet of data is ready to send. The destination send READY indications after CONNECT is asserted for four clock cycles. Each READY indication (at least four clock cycles long) indicates that the destination is ready to receive a burst (256 words of 32 bits each). The source will respond by asserting BURST, indicating that the burst is being transferred at a rate of one word per 25 MHz clock cycle. If the REQUEST line is dropped, the destination responds by releasing CONNECT, and the connection is broken. Normally this occurs after BURST and PACKET have been deasserted.

3.3.3 HIPPI Electronic MUX (Parallel--Serial PCB)

To achieve long haul fiber optic transmission of HIPPI data feasible, the parallel data, parity and control must be reduced to a single serial line. This is accomplished by a JPL designed HIPPI parallel-serial printed circuit board (PCB). The HIP PI parallel-serial PCB is composed of two major multiplex/demultiplex (MUX/DMUX) chip sets and their supporting components. The first MUX/ DMUX pair translates between the conventional parallel HIPPI format of 8 control and 32 data lines at 25 MHz to a 20 bit–wide form at 50 MHz. A second MUX/DMUX pair converts to/from the 20 parallel lines from/to a serial line which is clocked at a 1.2 GHz rate. MUX and DMUX operations are performed on the same multilayer PCB. An option is included for a local loopback mode in which serial output of the MUX is sent directly to the DMUX input (which is useful for verifying that the PCB is operating correctly).

The second MUX/DMUX pair is responsible for setting up the 1.2 Gbit/sec link before actual HIPPI data can be sent. Functionally, this MUX/DMUX pair contains a transmitter, receiver, and state machine [11]. The state machine controls the status of the link and has three possible states: frequency acquisition (state O), waiting for peer (state 1), and sending data (state 2). The state machine decides what state it should be in based on its rncmory and the type of frame currently being received: fill word O (FW0), fill word 1 low or high (FW 1L or FW lH), data/control word, or an error frame (i.e. the frarnc is invalid). When the state machine is in state O, it is in the reset state.

The transmit chip sends FWO continuously, and the receiver phase locked loop (PLL) is in frequency detection mode. When the receiver detects either FWO or FW1, the state machine is advanced to state 1. At this stage, the receive phase–locked loop (P] .L) is phase-locked and the transmit chip sends FW 1 (1./} 1). If the receiver detects FW0, it remains in state 1. If FW1 or a data word is detected, the state machine advances to state 2. Now data tl ansmission and reception are enabled for the parallel interface. This will result in the local HIPPI interconnect lines being asserted. When all HIPPI interconnect lines are asserted (which means the local and remote state machines are both in state 2), the link has been established and HIPPI data communications can proceed, A state machine diagram appears in Fig. 21.

Figure 22 shows a sample oscilloscope trace of the laser output of the serial HIPPI data stream. In this case, the link is transmitting a stable data word that corresponds to the link being established, but no actual data being transferred. In order to determine the minimum laser current bias required for error–free transmission, the bias current was varied over a range from around threshold to 60 mA, and the corresponding bit error rate (BER) measured (for random data pattern). The power at the receiver was fixed at –lo dBm. The result in Fig. 23 indicate that for current biases below 42 mA, the BER is markedly increased due to relaxation oscillation like that shown previously in Fig. 15. Figure 24 shows the results of a test in which the laser current bias was fixed at 45 mA and the received optical power varied via attenuation in the fiber link. It can be seen that for a received optical power greater than approximately –20.7 dBm the BER is at least 1()-^{11}.

For a BER of 10-1], this link configuration gives the requirement that the power available to the receiver be greater than –20.7 dBm. Figure shows the link power budget analysis for the final link using the leased telecom line. This link is approximately 14 Km in length, but has many lossy splices with an estimated loss of 10dB. This fact, combined with coupler and laser--to-fiber coupling losses, necessitate the use of an optical amplifier to boost the signal up to a level the receiver can detect.

4. Applications

The Global Grid Data Fusion program looks to develop and appl y state of the art information system technologies to the *real-time* remote theatre defense. Specifically, key technologies such as all-optic terabit networks and teraflop parallel supercomputers, will be used to collect, analyze and correlate multiple image sensor inputs from the battlefield theatre via the Global Grid, locate and classify targets (either ground or aerial), assess defensive strategies through fast battlefield simulators, and finally project solutions back to field commanders in the form of high resolution 3D terrain visualizations. By definition, the response time of this complete system must be less than the reaction time of the weapons systems at the remote theatre, typically a few seconds or less,

4.1 Background

Global Grid represents a new Department of Defense operational doctrine. This doctrine postulates a defense future in which DoD people and resources are highly inter-connected and shared, often by non-DoD users. Past distinctions between tactical and strategic, local and global, indeed defense and non-defense break down and become blurred. Mission planning and situation assessment via teleconferencing become global, tactical real-time military actions arc backed up by computing and analysis capabilities remote from the field of action, sensor platforms are used to support disaster relief as well as for mobile launcher detection.

Lessons learned from the scud missile strikes during the Gulf War and the more recent humanitarian relief efforts in Africa, clearly indicate satellite imagery, if analyzed and understood *quickly* enough, can reverse a military tactical advantage, or in the case of disaster mitigation, stabilize or even reverse a disease epidemic. Such high resolution satellite imagery (e. g., radar, infrared, or hyperspectral) is typically widely dispersed geographically today, making it difficult to collect quickly. Furthermore, the computational resources needed to process it in real-time often involve large highly specialized supercomputers at various national laboratories. Typically, these already existing resources are not used during crisis management because of inadequate communications infrastructure.

The Global Grid Data Fusion Testbed will apply information and communication system technologies in the following areas: TeraFLOP MPP supercomputer technologies, high-speed low-latency fiber optic networks (100 Gbit/s – 1 Tbit/s), WAN based meta-supercomputers, advanced data base management (DBM) and battlefield management (BM) simulators, multi-spectral satellite and ground-based imagers and radarGigabit relay satellites, neural and fuzzy logic tactical situation assessment algorithms, force structure characterization, force elements geolocation/prediction, and tactical intelligence knowledge engineering (Intelligence Template Mappers).

4.2 Andromeda Project

The *Andromeda Project* proposes to glue together these high performance data archive and processing centers with a high performance multi-gigabit network to create a nation-wide *n2eta-supercom-*

puter that can be tapped into or exported to distant continents by the Global Grid networks. Much like the civilian air transport reserve was used to transport troops during the Gulf crisis, this vast network of data imagery and highest possible performance supercomputers, always being updated, could be applied on a moments notice in a global crisis. This would provide commanders in the field rapid (in minutes) analysis rapidly unfolding crisis, such as population migration, whereabouts of scud missile launchers, spread of brush wildfires, and flood prediction to name but a few.

An illustration of how CON US supercomputer assets might be projected to remote corners of the world via the global grid is shown in 27. Here, remote sensor data first is collected on a remote theatre of operations. This includes on-site sensor data (GPS, radar, IR, IMINT, ELINT, and others) delivered over tactical battlefield C^3 networks as well as satellite image data transmitted back to CONUS over traditional SATCOM channels. Remote ground sensor data, concentrated, and then transmitted via a gigabit relay satellite back to CON US where a national meta-supercomputer network processes the data (noise removal, map registration, etc), and then performs the *data fusion* operation it self---multi-modality correlations between image data sets and existing data bases. Targets and positional coordinates are then supplied to sophisticated battlefield managers and simulators (also meta-supercomputer based) that deliver tactical planning data in the form of 3D visualizations back to field commanders over the return path of the Global Grid network—--all in a few seconds. It is our goal to make the entire response time short enough that it might even be able to operate as machine in the loop, as part of, for example, a Patriot advanced early warning guidance system, or in providing target location data to F15 fighter pilots during the progress of missions.

An underlying premise of this effort is that some of the nighest performance computation and communication systems being developed today can be used to transport, process, and disseminate remote theatre sensor image and radar data *as fast as* it is being collected, Specific technology examples include: massively parallel processor (MPP) supercomputers to process C^3 data in real-time, ubiquitous acquisition and dissemination of large data sets via the Global Grid, creation of powerful, highly reliable and available *meta-supercomputer* networks within CONUS with all-optic terabit networks. In brief, this effort will provide an experimental testbed to explore high performance systerns technologies for next generation *WarBreaker* and the resulting new enabling applications for theatre defense.

5. Future Work

Future work focuses on continued development of the components and subsystems discussed above. Specifically, the areas of interest arc construction of an 8-node MCSN testbed, addressing issues related to using an optical data plane with the electronic MCSN node for optical packet routing, completion and demonstration of the 4-wavelength WDM HIPPI link, and integration of the WDM laser array with a fiber star coupler,

5.1. MCSN Development

The prototype MCSN node design provides the foundation for an 8-node MCSN demonstration system. The current switching node design will be converted to a PCB version to allow for ease of replication of the node. The purpose of this system is to validate the simulation results presented in Ref. [1]. Additional work to build the MCSN demonstrator system is needed to replace the on board packet generator shown in Fig. 7, with a more sophisticated off board generator to allow for different types of traffic patterns and collection of routing statistics. This generator is currently envisioned as a peripheral card which can be plugged into a PC. With this configuration, we can network up to eight PCs for system performance evaluation in addition to realizing a generalized interconnection for parallel computing applications.

While the MCSN demonstrator system uses electronic packets for rapid system evaluation, this system can also route optical packets due to the short packet deflection protocol described in Ref. [1]. To realize a multi-gigabit/see system requires replacing, the electronic data plane in Fig. 3 with an optical equivalent. The primary limitation of current commercially available optical switches is the millisecond set uptime, which is not well suited for routing small packets with a deflection protocol. Integrated switching fabrics provide nano second switching speeds, but at expense of increased noise due to crosstalk and optical amplifier noise [13]–[20]. This technology is not yet mature enough for application to the data plane of an MCSN node but the implementations in Refs. [13], [14], [18], [19], *[20]* look quite promising and warrant fu1 ther investigation.

5.2. WDM Components

The main focus of the 4-wavelength WDM HIPPI link to this point has been the design and testing of the electronics for the parallel-serial HIPPI PCB, interfacing this board to the 4–element stepped wavelength array and characterization the array elements. The next task is to build the requisite number components to build the 4-wavelength link for full-duplex operation in a laboratory setting. This link will be tested to verify and characterize the operation of the four channels before installation of the link between the supercomputing facilities of JPL and the California Institute of Technology (CIT). The purpose of the laboratory testis to validate link operation before introducing additional variables due to transmission line characteristics of the 10 Km link between the two sites.

The primary difficulty in using the MDL laser arrays comes from the labor intensive approach needed to couple the four laser elements into a single fiber. The alignment of the fibers to the array requires three-dimensional positioning and results in low, non-uniform coupling efficiencies. To alleviate this problem, a monolithically integrated laser array and star coupler are presently being developed by the MDL to improve the overall laser-fiber coupling loss. This WDM transmitter will include four DFB lasers and a 4x 1 star coupler, thus requiring only a single fiber pigtail and potentially reducing coupling loss and coupling non-uniformity across the array. This star coupler-based monolithically integrated WDM transmitter is also scalable to an N-element laser transmitter with an integrated *N*x1star coupler–a21 -element monolithically inte.grated transmitter has already been demonstrated that requires coupling to a single fiber [12].

6. Conclusions

Significant progress has been made in defining the basic components and subsystems needed to realize an all-optical data path network. A network architecture called a multi-cylinder shifflenet was conceptualized and simulated to verify datagram routing of asynchronous optical packets. This simulation model incorporated PE switching nodes to handle short, asynchronous packets and an interface to convert long, streaming packets into short packets well suited to the MCSN concept with a WDM data format.

As a complement to the theoretical and simulation work done, various components and subsystems are under development to validate the proposed WDM networking concept. These efforts include a prototype MCSN switching node for routing asynchronous datagram packets, a HIPPI serial-parallel PCB as the first step in realizing the desired WDM network interface, and the 4-element stepped wavelength laser arrays. The last two elements are being integrated into a four channel HIPPI extender to ultimately demonstrate the viability of WDM.

These components and subsystems are viewed as intermediate steps to the final optical network as an enabling technology to multi-gigabit networking applications. Additional progress in both the size of optical switching fabrics and the data path characteristics is needed before these devices can be inserted into the proposed network. We further look to WDM techniques to utilize the THz bandwidth of optical fiber. This technology, however, requires further integration of the WDM laser array (and a detector array) with an integrated optic coupler to simplify system construction and mitigate coupling losses. Encouraged by the current results, we can begin to see the formation for a true multi-gigabit/sec interconnection network.

7 Acknowledgements

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Ballistic Missile Defense Organization, Innovative Science and Technology Office, and the National Aeronautics and Space Administration. The work was performed as part of JPL's Center for Space Microelectronics Technology.

Reference herein to any specific commercial product, process, or service. by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government, the Jet Propulsion Laboratory, or the California Institute of Technology.

8 References

[1] S. P, Monacos, and A. A. Sawchuk, *“A Scalable Recirculating Shuffle Network with Deflection Routing*”, submitted to the Special Joint Issue of IEEE J. on Selected areas in Communications and IEEE J. of Lightwave Technology on OPTICAL NETWORKS.

[2] S. P. Monacos and A. A. Sawchuk, *“A Permutation Engine Switching Node”,* submitted to the J. of Parallel and Distributed Computing.

[3] J. R. Sauer, *“An Optoelectronic Multi–Gb/s Packet Switching Network*,” *OCS Technical Report 89-06,* February 1989.

[4] N. F. Maxemchuk, *“Comparison of Deflection and Stcjre-arid-l;or\i’{lrd Techniques in the Manhattan Street and Shuffle-Exchange Net works,”* IEEE INFOCOM *’89,* pp. 800-809, April 1989.

[5] A. K. Gupta and N. D. Georganas, *”Analysis of a Packet Switch with Input and Output Buffers and Speed Constraints,”* IEEE INFOCOM ‘91, Pp. 694-700, 1991.

[6] M. Aoki, T. Tsuchiya, K. Nakahara, M. Komori and K. Uomi,*”High- Power and Wide–Temperature–Range Operations of In GaAsP–InP Strained MQW Lasers with Reverse–Mesa Ridge Waveguide Structure*,” Photonics Technology Letters, vol. 7, no. 1, pp. 13--15, 1995.

[7] T. Koch, *”Lasers Sources for Wavelength Division Multiplexing,’*’ Conference on Optical Fiber Communications, paper WF, Feb. 1995.

[8] E.E Bert Basch, cd., *Optical Fiber Transmission,* Howard W. Sams & Co., p 309-313, 1987.

[9] T. Russel, *”HIPPI and the issues of HIPPI Data Net working* ,“ Ultra Network Technologies, 1991.

[10] Draft AMERICAN NATIONAL STANDARD X3.183--199x

[11] *”Preliminary Specification for the Gigabit Rate Transmit Receive Chip Set”,* Hewlett Packard, 1992.

[12] C. E. Zah and T. 1? Lee, *“Monolithically Integrated Multi- Wavelength DFB Laser Arrays and Star Couplers for WDM Light wave Systems”,* Optics and Photonics News, pp. 24–27, March 1993.

[13] W. H. Nelson, A. N. M. Masum Choudhury, M. Abdalla, R. Bryant, E. Meland, W. Niland, and W. Powazinik, "*Large-angle* 1,3pm *InP/InGaAsP digital optical switches with extincion ratios exceeding 20 dB*," OFC 1994 Technical Digest Series, pp. 53--54.

[14] T. Kirihara, M. Ogawa, S. Tsuji, and H. Inoue, *"i]igh-speed signal-transmission performance in a lossless 4x4 optical switch for photonic switch ing,"* OFC 1994 Technical Digest Series, pg. 55.

[15] T. Kirihara, M. Ogawa, H. Inoue, and K. Ishida, "*Lossless and Low- Crosstalk Characteristics in an InP -Based 2x2 Optical Switch,"* IEEE Photonics Technology Letters, vol. 5, no. 9, pp. 1059–1 061, September 1993.

[16] P. Granestrand, B. Langerstrom, P. Svensson, H. Olofsson, J. E. Falk, and B. Stoltz, *"Pigtailed Tree-Structured 8x8 LiNbO3 Switch Matrix with 112 Digits] Optical Switches,"* IEEE Photonics Technology Letters, vol. 6, no. 1, pp. 71-73, January 1994.

[17] W. H. Nelson, private communication.

[18] T, Kirihara, M. Ogawa, H. Inoue, H. Kodera, and K. Ishida, "*Lossless and Low-Crosstalk Characteristics in an InP-Based 4x4 Optical Switch with Integrated Single-Stage Optical Amplifiers,"* IEEE Photonics Technology Letters, vol. 6, no. 2, pp. 218--221, February 1994.

[19] R. Nagase, A. Himeno, M. Okuno, K. Kate, K. Yukimatsu, and M. Kawachi, "*Silica-Based 8x8 Optical Matrix Switch Module with Hybrid Integrated Driving Circuits and its System Application,"* J. of Lightwave Technology, vol. 12, no. 9, pp. 1631--1639, September 1994.

[20] W. H. Nelson, A. N. M. Masum Choudhury, M. Abdalla, R. Bryant, E, Meland, and W. Nil and, "Wavelength- *and Polarization-Independent Large Angle InP/InGaAsP Digital Optical Switches with Extinction Ratios Exceeding 20 dB*," IEEE Photonics Technology Letters, vol. 6, no. 11, pp. 1332– 1334, November 1994.

List of Figures

Figure 19 Parallel HIPPI connections.

Figure 20 HIPPI signaling sequence.

Figure 21 State machine diagram.

Figure 22 Optical serial HIPPI 'rest' pattern.

Figure 23 BER vs laser current.

Figure 24 BER vs loss characteristic.

Figure 25 Link power budget analysis.

Figure 26 V/I laser characteristic.

Figure 27 In a typical Global Grid Data Fusion application, CONUS assets are projected to a remote theatre. First, remote sensor data is collected on a remote theatre of operations, concentrated, and transmitted via a gigabit relay satellite to CONUS where a national meta-supercomputer network processes the data, performs multi-modality correlations between image data sets and existing data bases. Targets and positional coordinates are then supplied to sophisticated battlefield managers and simulators (also meta-supercomputer based) that deliver tactical planning data in the form of 3D visualizations back to field commanders over the return path of the Global Grid network—all in a few seconds.

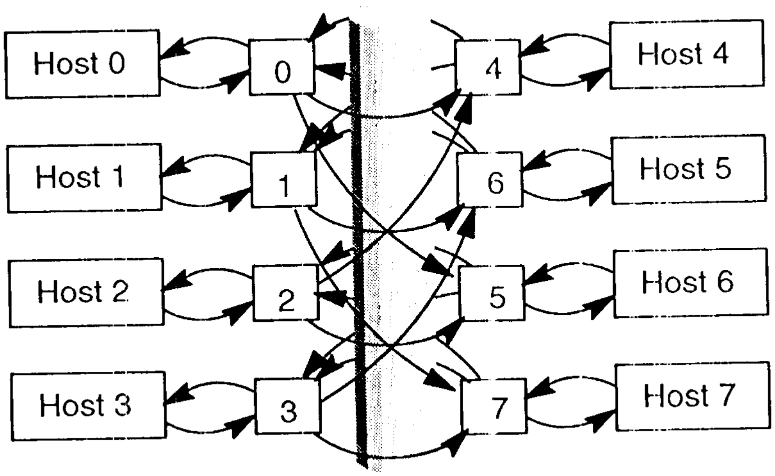


Figure 1

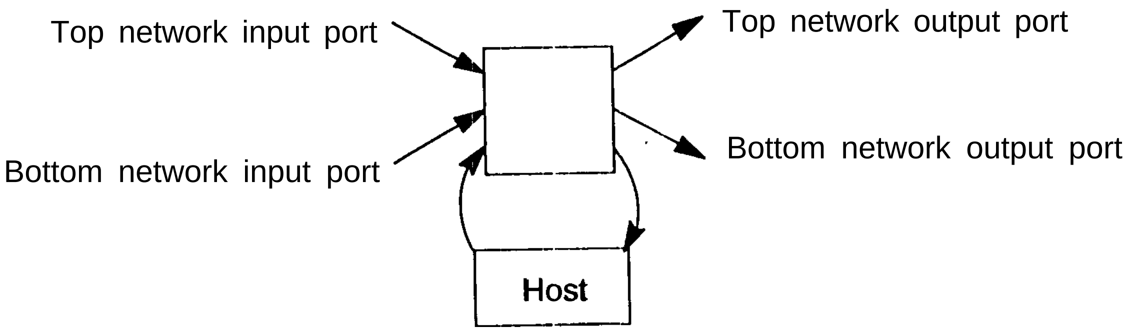


Figure 2

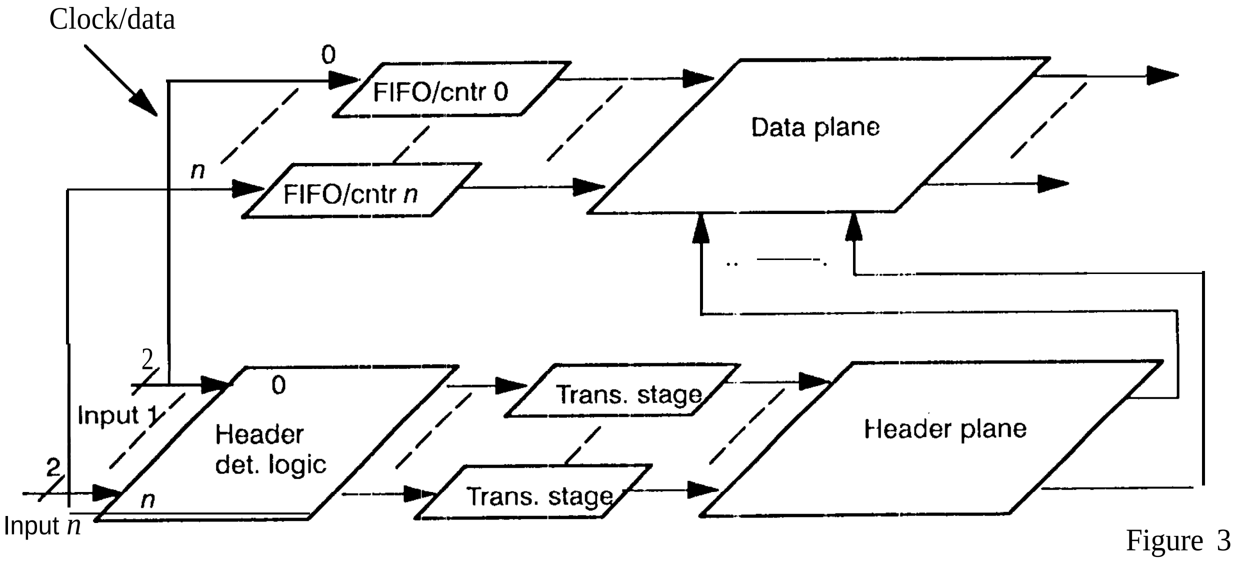

Clock/data
0
FIFO/cntr 0
n
FIFO/cntr n
Data plane
2
0
Input 1
Header
det. logic
2
n
Input n
Trans. stage
Trans. stage
Header plane

Figure 3

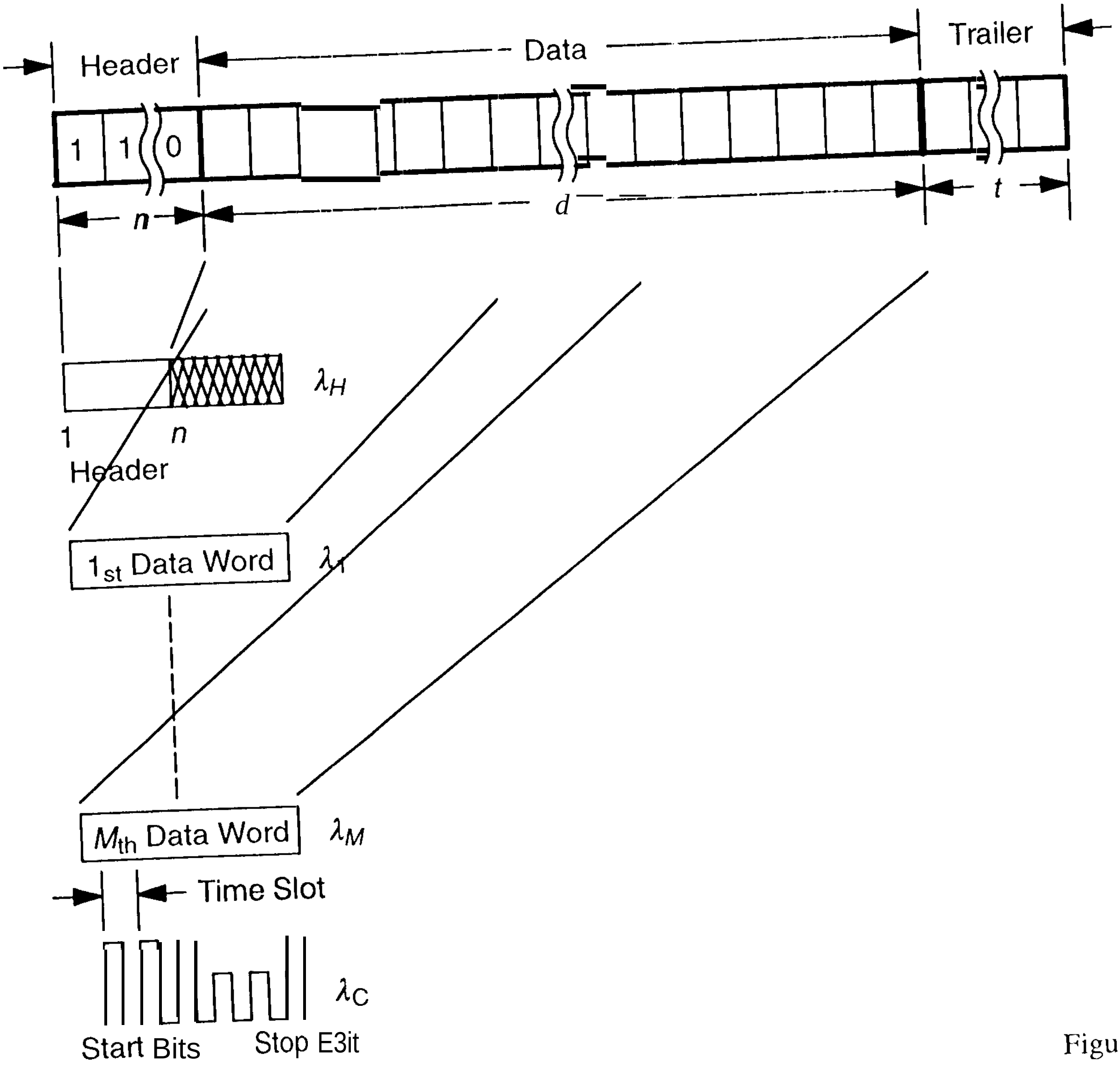

Header
Data
Trailer
1
1
0
n
d
t
λ_H
1
n
Header
1_{st} Data Word
λ_1
M_{th} Data Word
λ_M
Time Slot
λ_C
Start Bits
Stop E3it

Figure 4

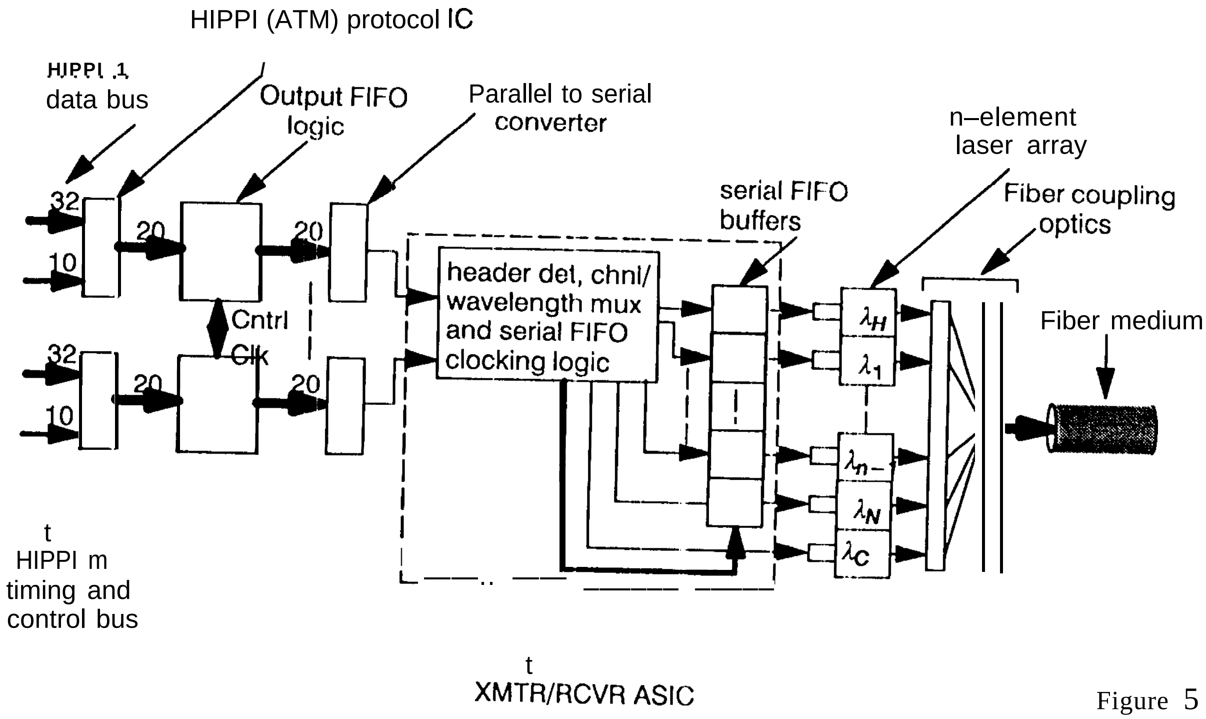


Figure 5

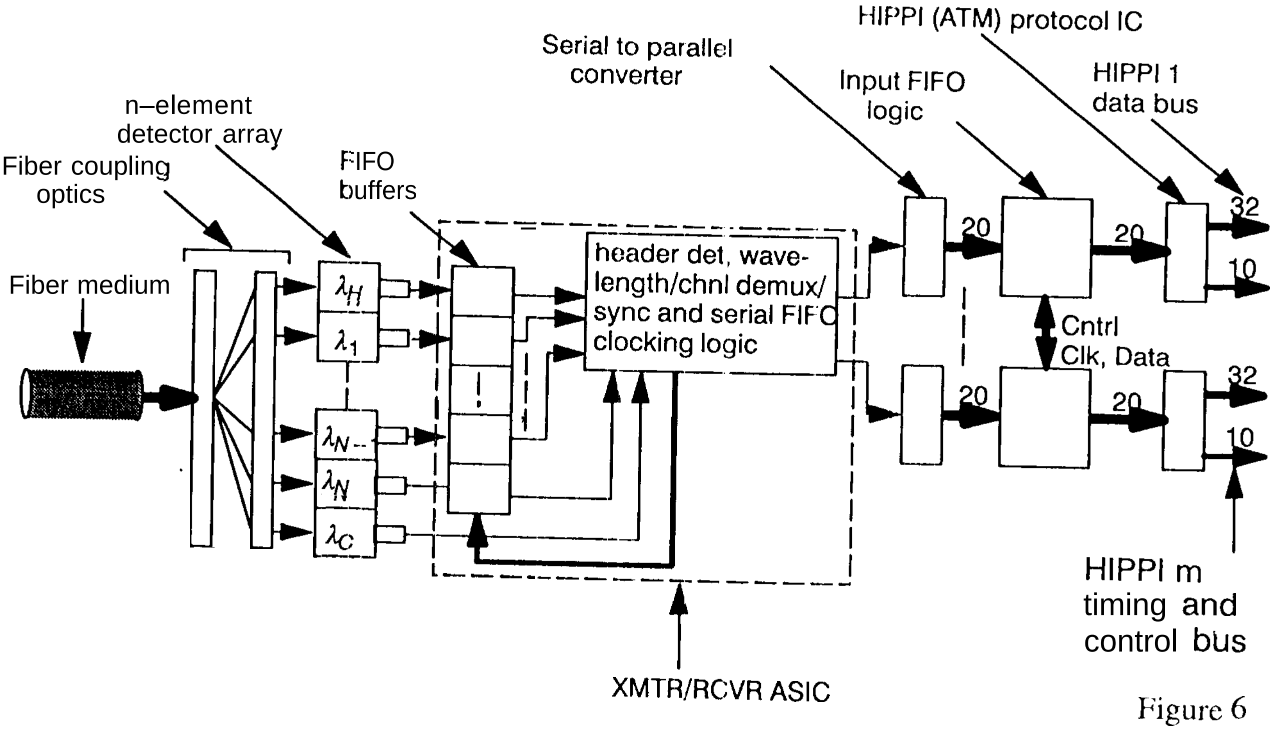


Figure 6

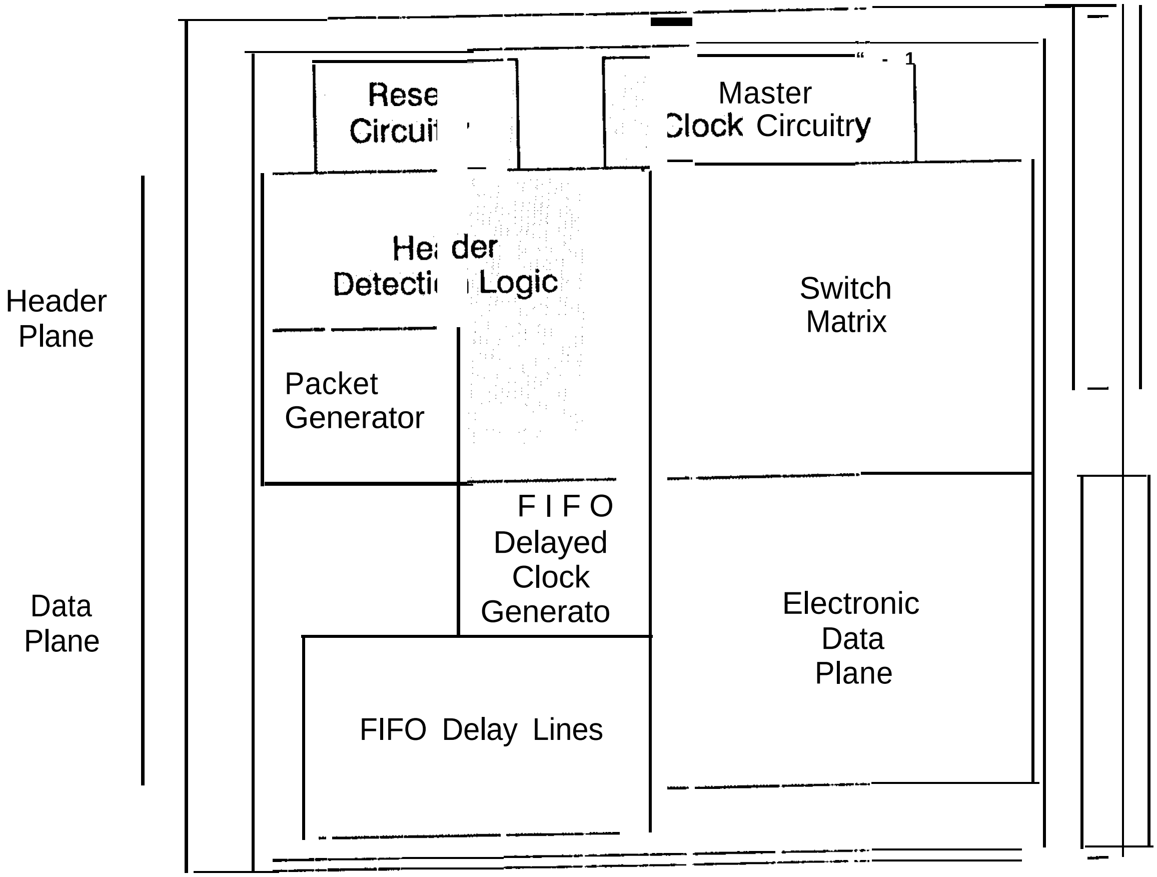

Header
Plane
Data
Plane
Rese
Circui
Master
Clock Circuitry
Hea der
Detecti Logic
Switch
Matrix
Packet
Generator
F I F O
Delayed
Clock
Generato
Electronic
Data
Plane
FIFO Delay Lines

Figure 7

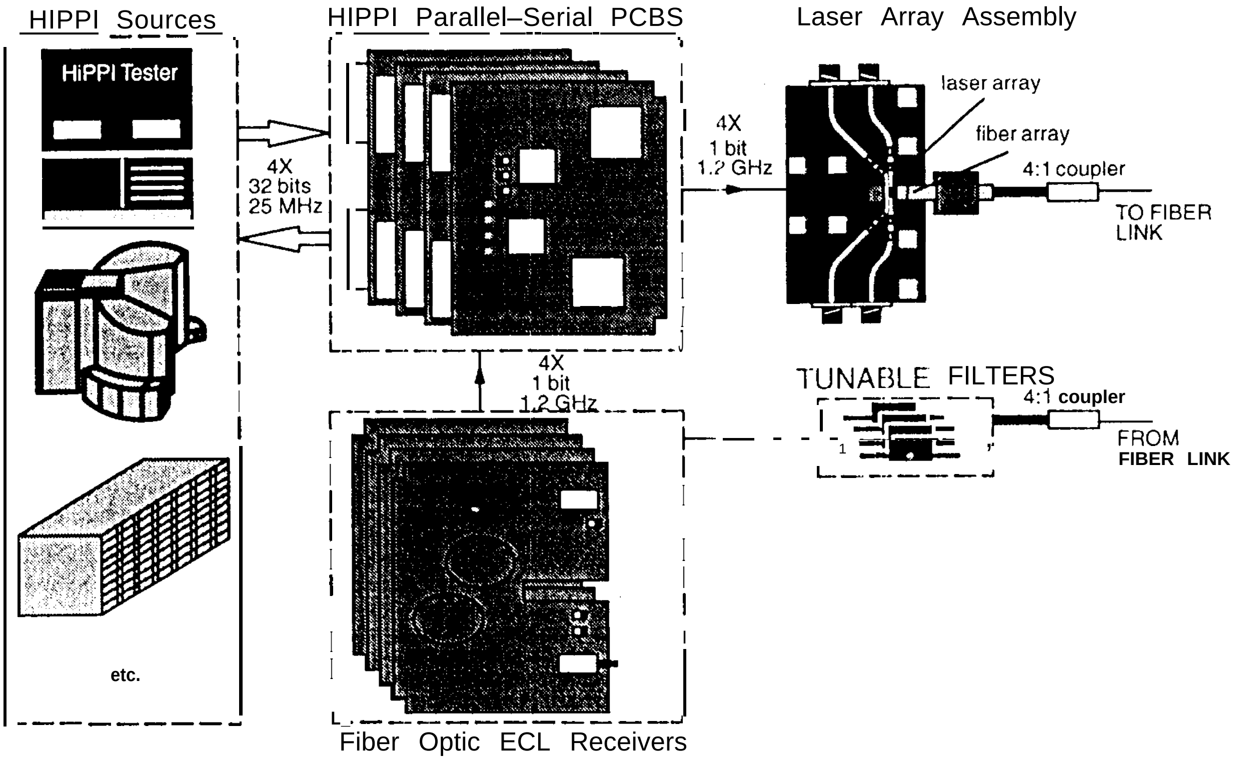


Figure 8

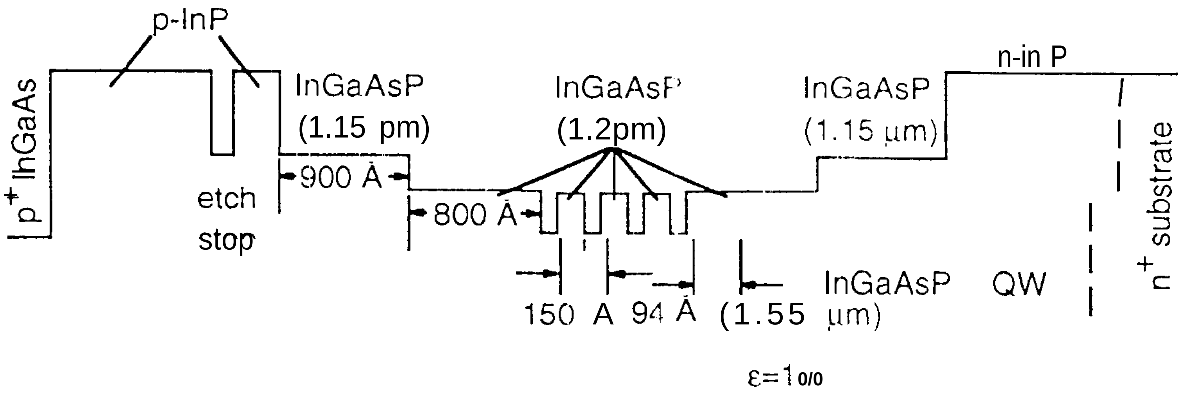


Figure 9

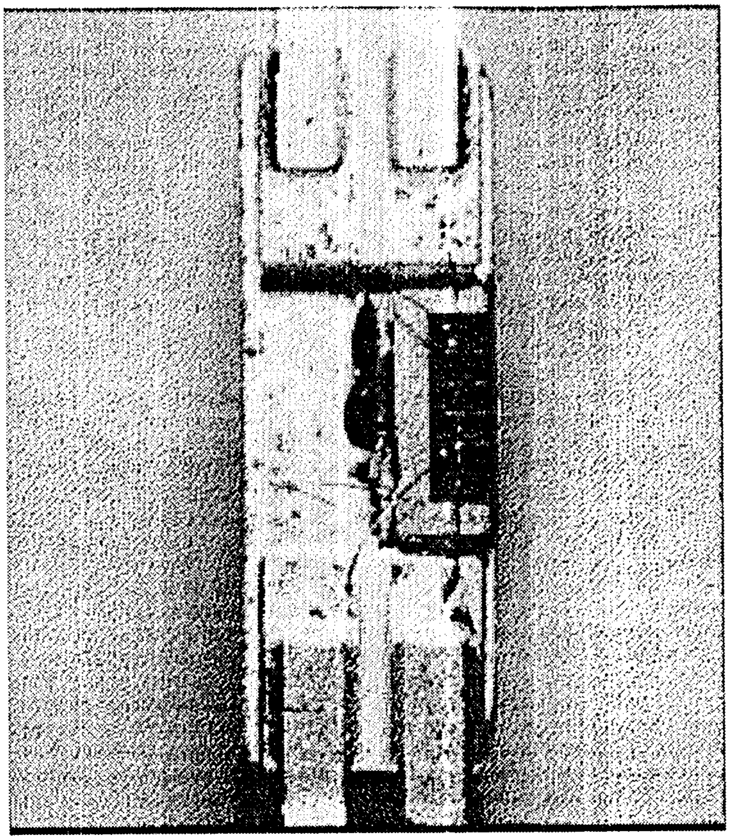


Figure 10

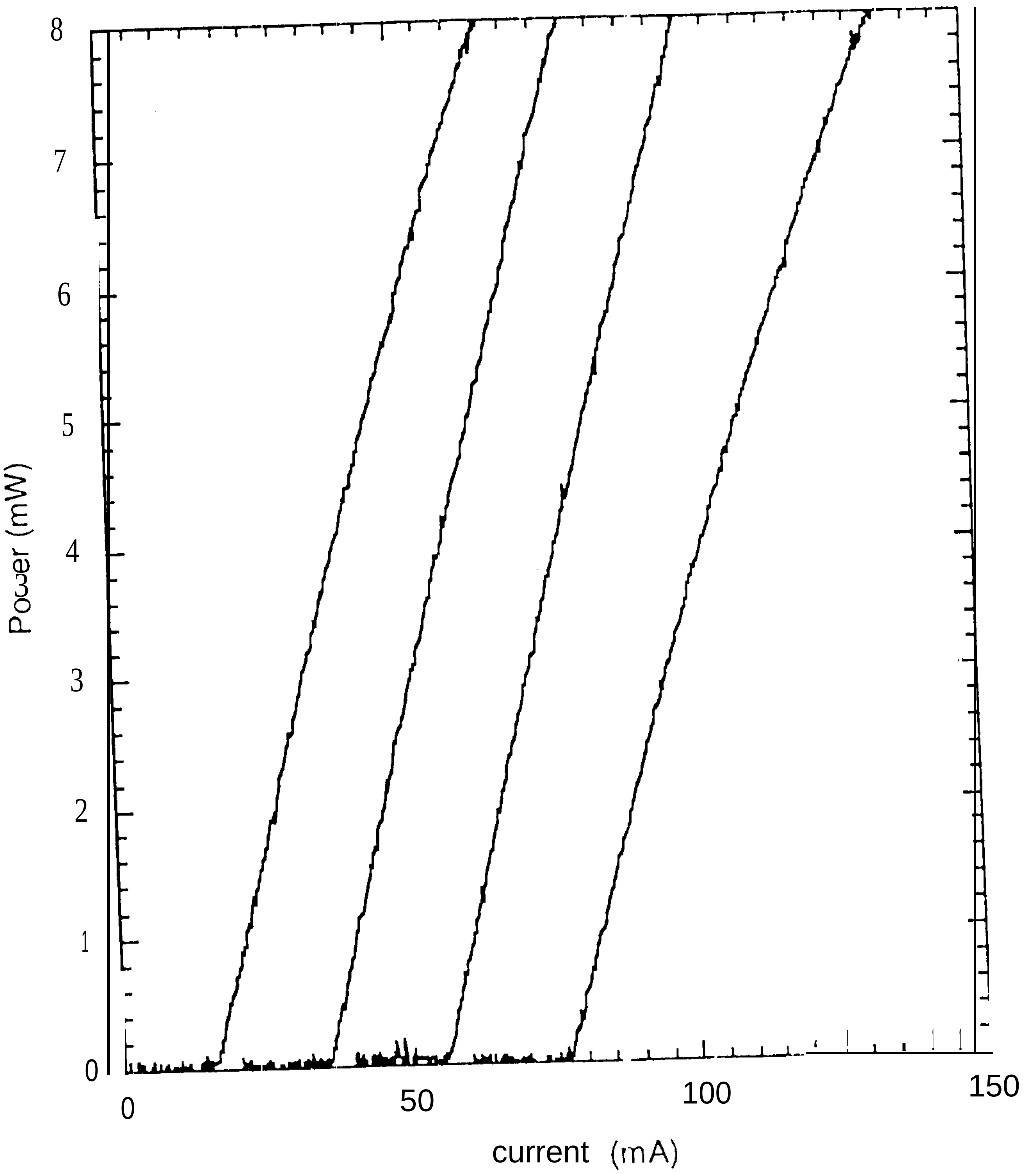


Figure 11(a)

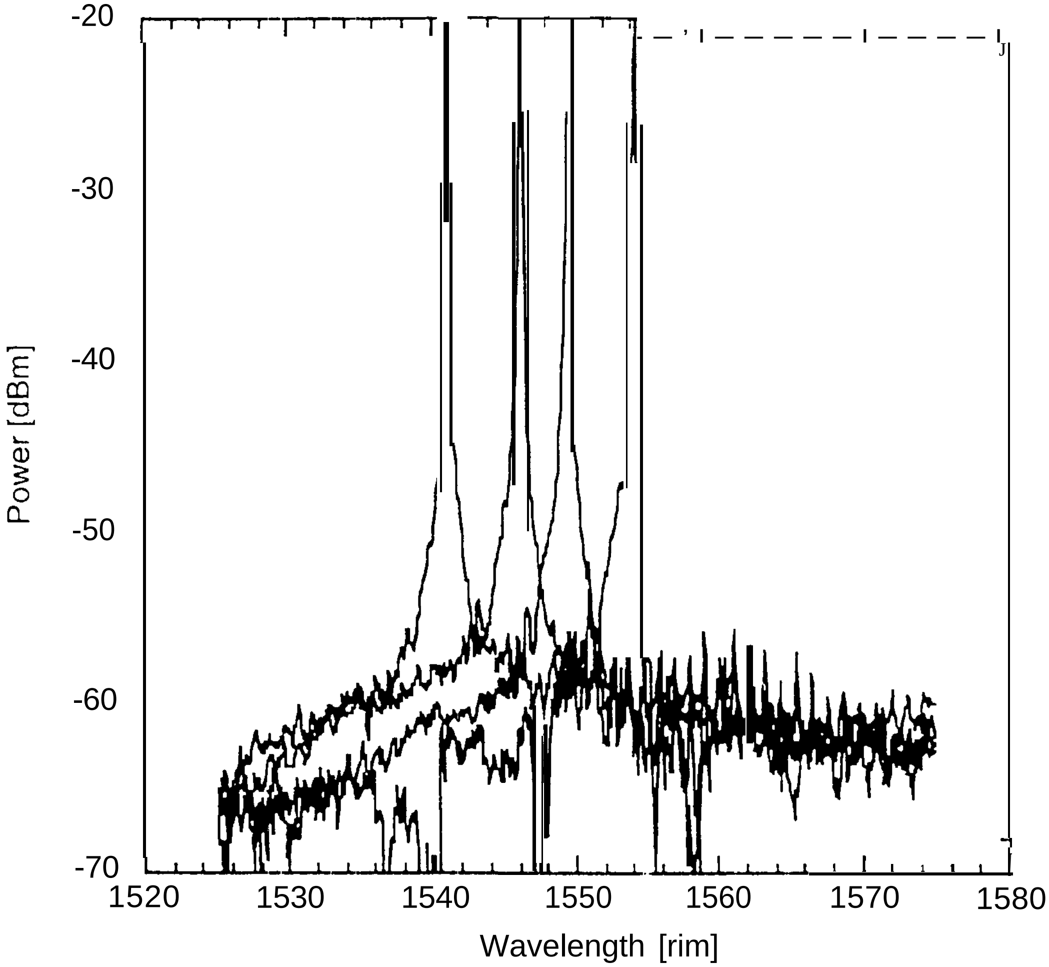

-20
-30
-40
-50
-60
-70
Power [dBm]
1520
1530
1540
1550
1560
1570
1580
Wavelength [rim]

Figure 11(b)

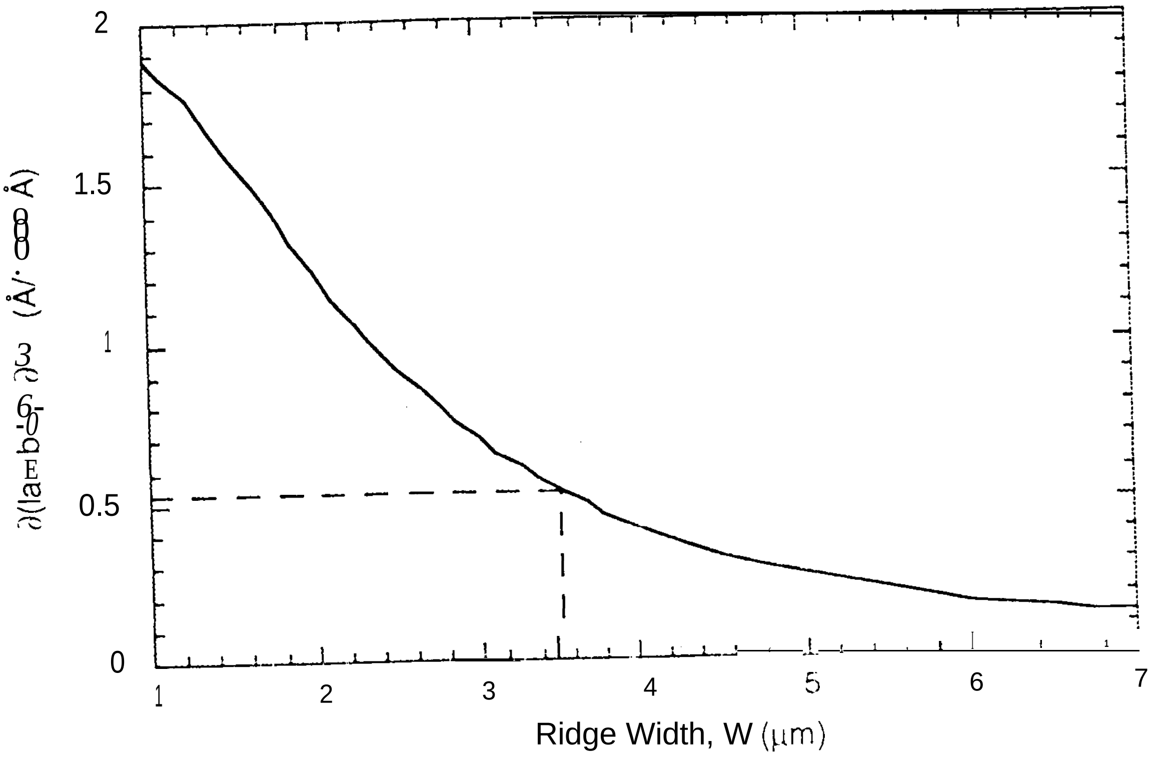


Figure 12

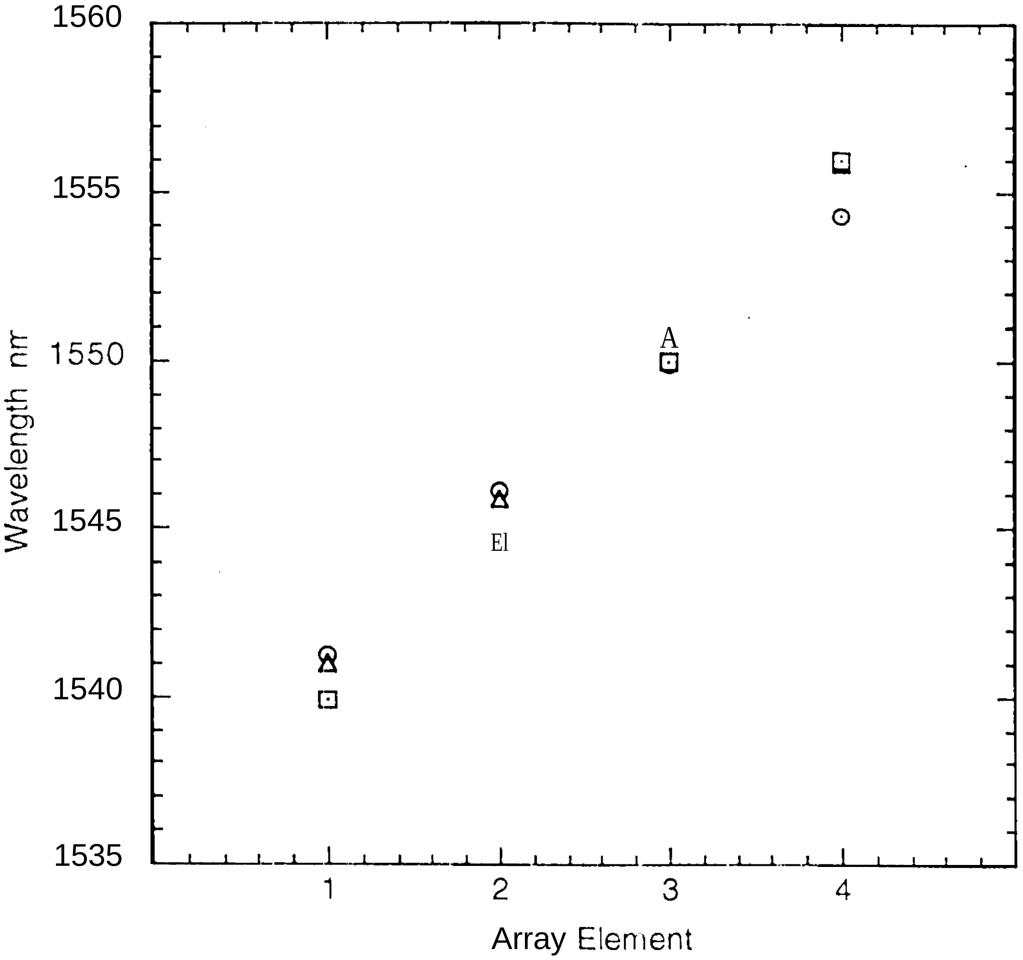

1560
1555
1550
1545
1540
1535
Wavelength nm
A
El
1
2
3
4
Array Element

Figure 13

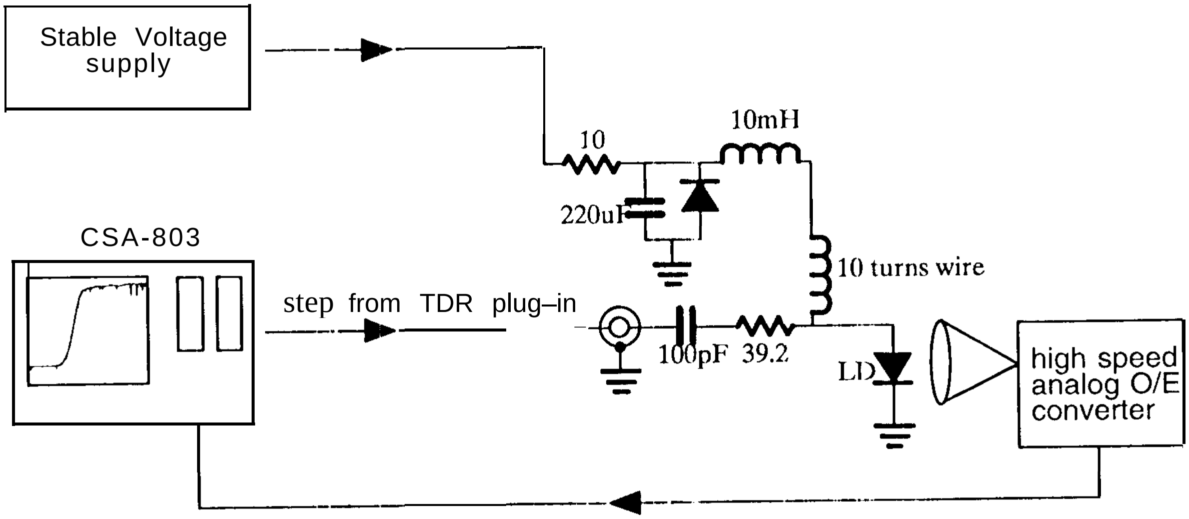

Stable Voltage supply
10
10mH
220uF
CSA-803
10 turns wire
step from TDR plug–in
100pF
39.2
LD
high speed analog O/E converter

Figure 14

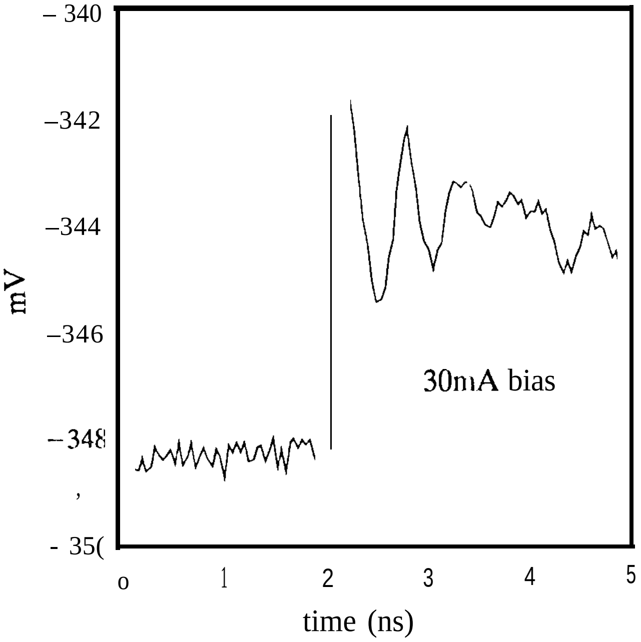


Figure 15

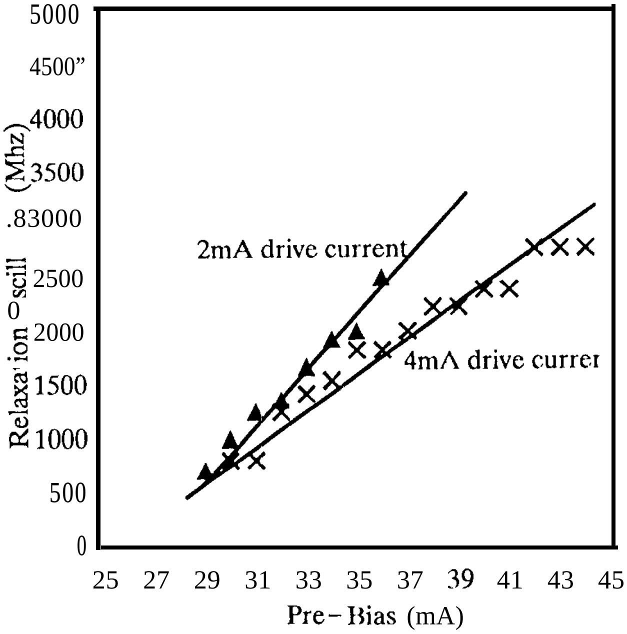


Figure 16

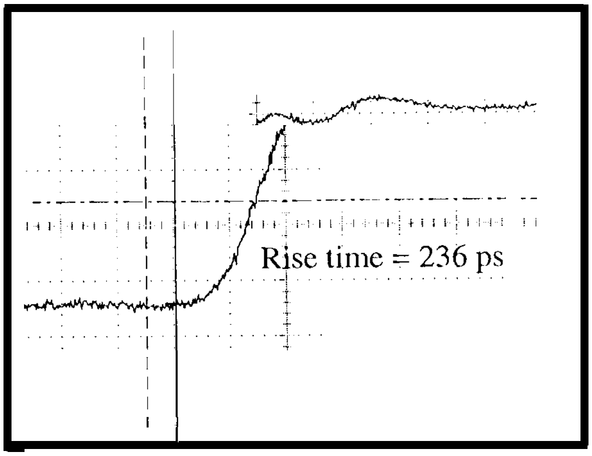


Figure 17a

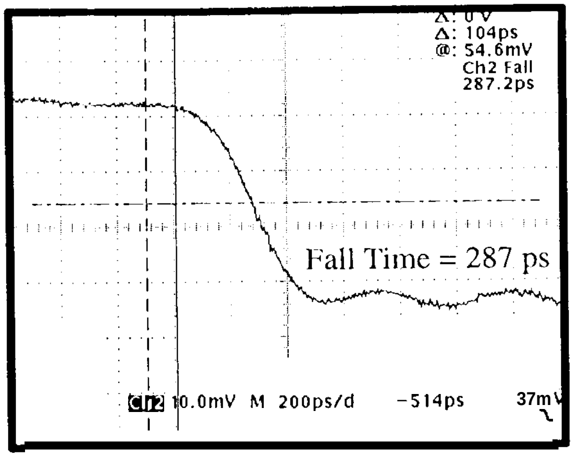


Figure 17b

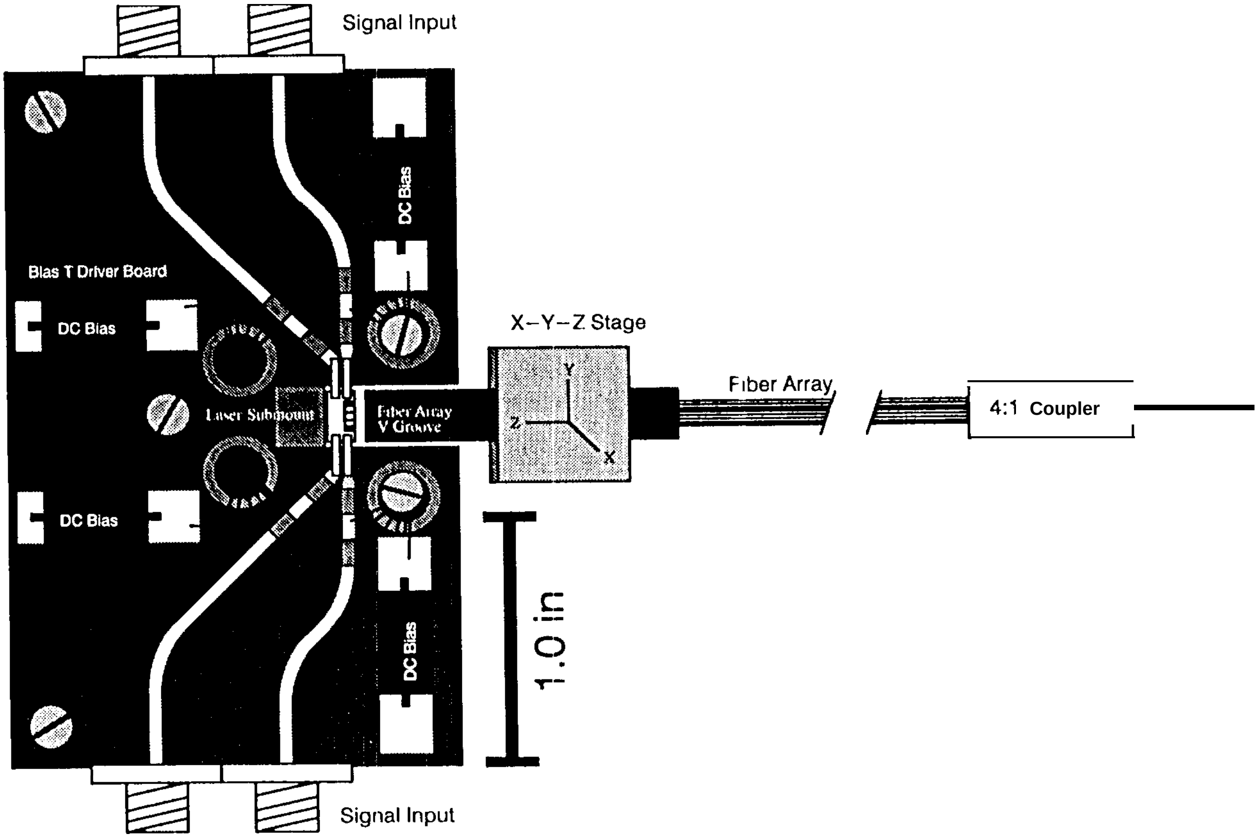


Figure 18

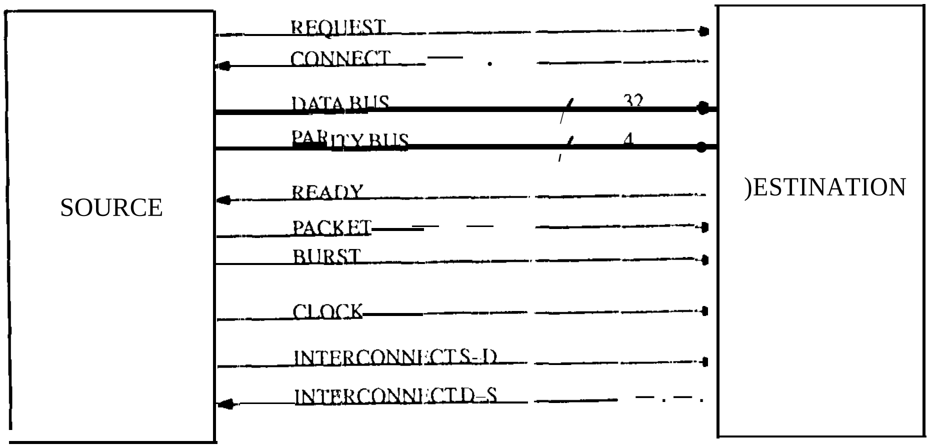


Figure 19

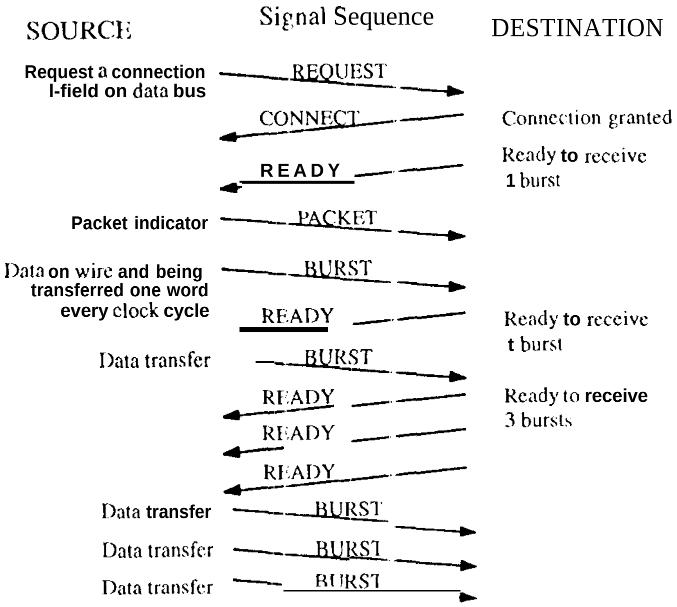


Figure 20

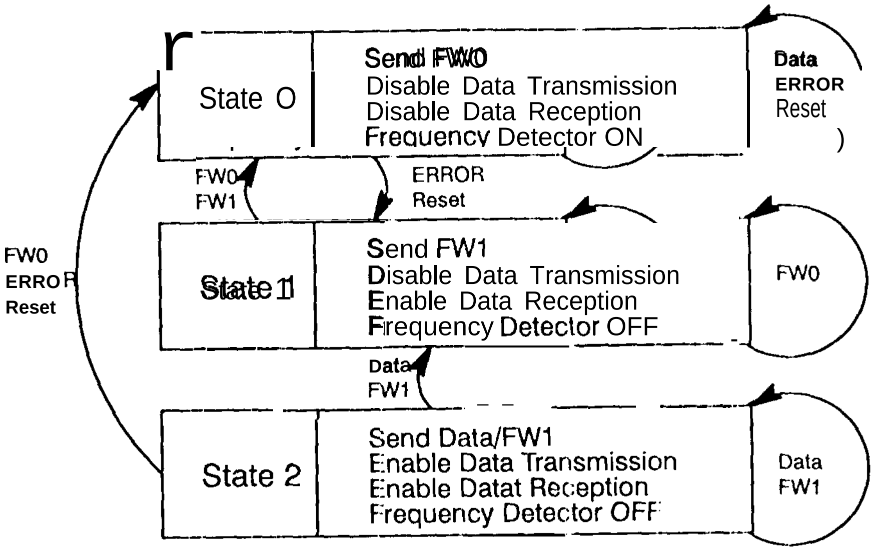


Figure 21

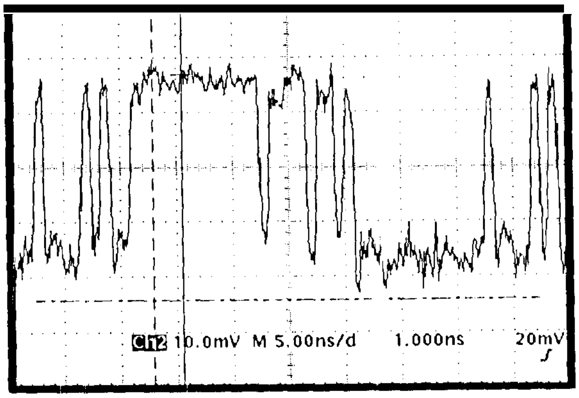


Figure 22

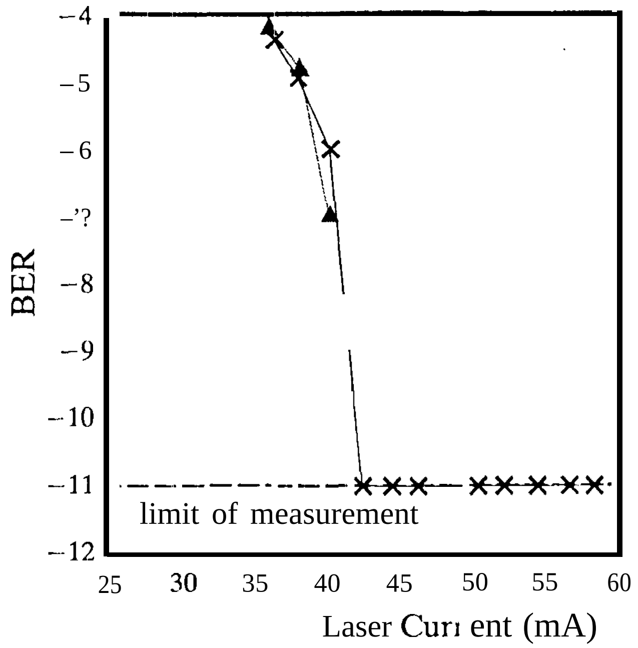


Figure 23

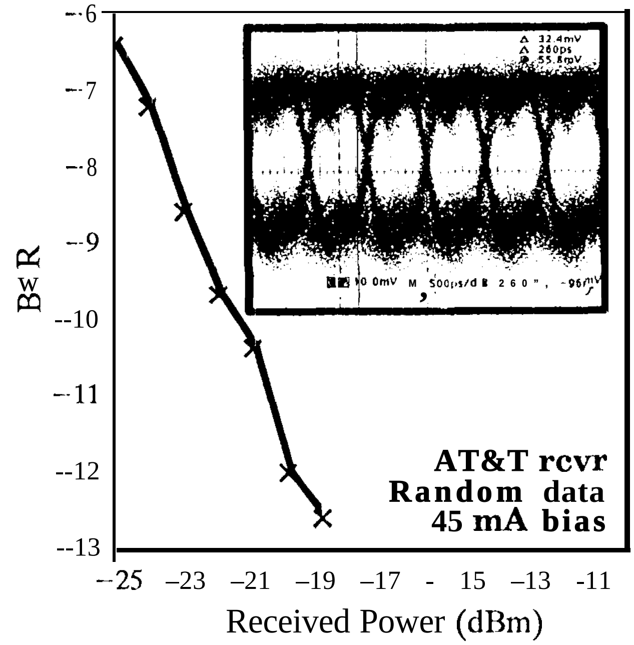
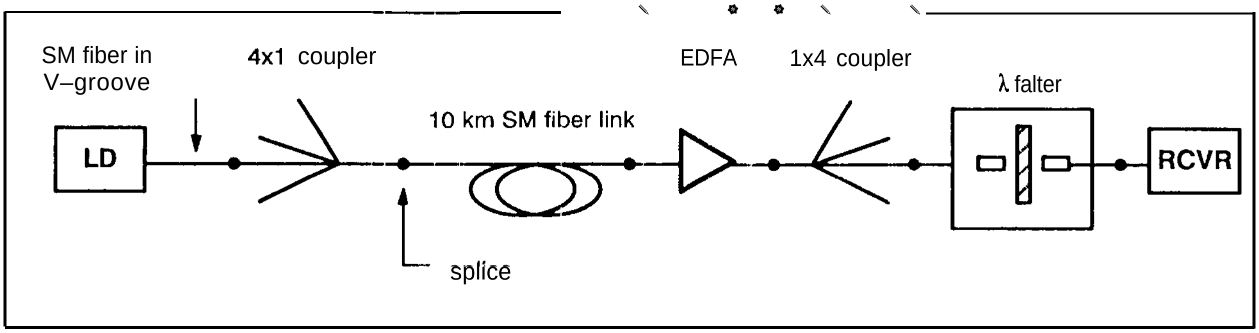


Figure 24



1.OSS Mechanism	Attenuation	Optical Power
LD to fiber coupling loss		−6.0dBm
4x1 coupler loss	6.0 dB	−12.0dBm
Leased Fiber Link Loss	10.0 dB	−22.0dBm
Erbium Doped Fiber Amplifier (EDFA)	−25dB	+3.0dBm
Splice losses (worst case)	5.0dB	−2.0dBm
1x4 coupler loss	6.0dB	−8.0dBm
Wavelength filter loss	3.5 dB	−11.5dBm

Figure 25

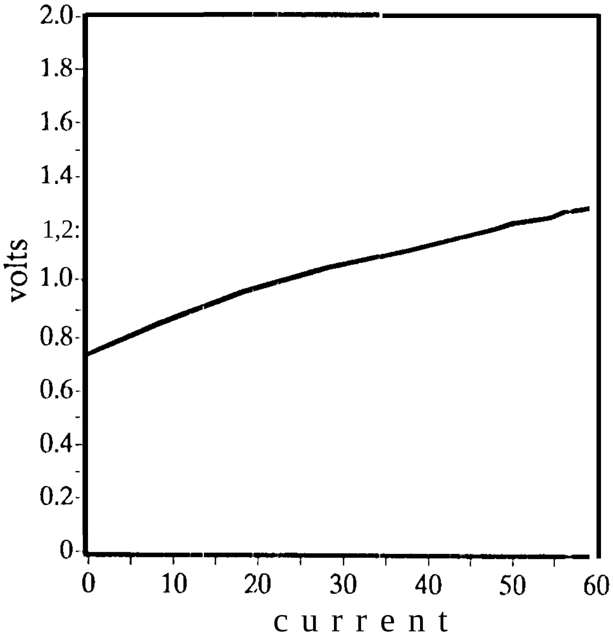
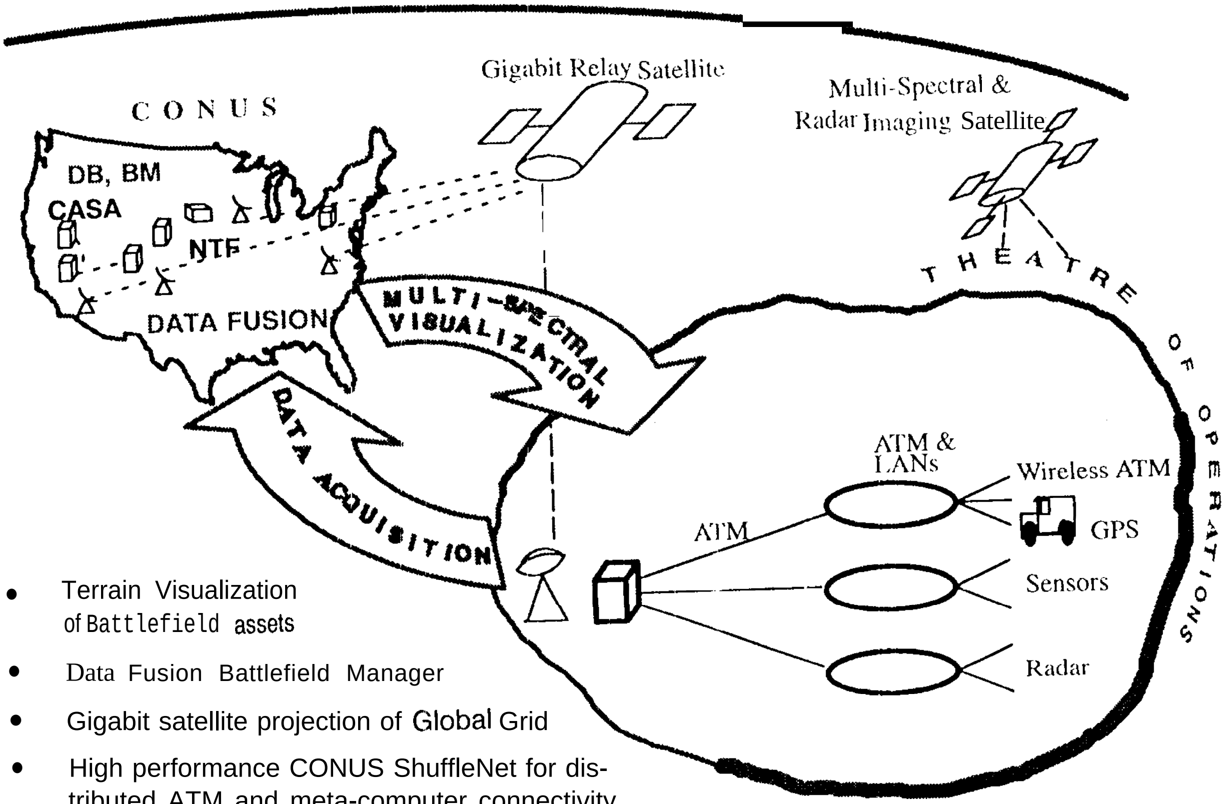


Figure **26**

GLOBAL GRID DATA FUSION APPLICATION



- Terrain Visualization of Battlefield assets
- Data Fusion Battlefield Manager
- Gigabit satellite projection of Global Grid
- High performance CONUS ShuffleNet for distributed ATM and meta-computer connectivity

Figure 27